Coarse Lexical Semantic Annotation with Supersenses

An Arabic Case Study

Nathan Schneider, Behrang Mohit, Kemal Oflazer, Noah A. Smith
Carnegie Mellon University, {nathan, behrang, kemal, noah}@cs.cmu.edu
Doha, Qatar

Common Confusions

<table>
<thead>
<tr>
<th>COMMUNICATON vs. COGNITION</th>
<th>ACT vs. COGNITION measures</th>
<th>Cognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT, LOCATION, ARTIFACT</td>
<td>ARTIFACT vs. COMMUNICATION</td>
<td></td>
</tr>
</tbody>
</table>

Boundaries

Fez, Morocco (1 or 2 LOCATIONS?)
World Cup Football Championship (single EVENT, or Football as an ACT)

Inter-annotator Agreement

Mentions: 70% F1, In-mention tokens: 71% rate, 83% F1, Cohen’s κ = .69

Training

Iterative annotation of 4 pilot articles and guideline refinement in consultation with annotators until their agreement reached an F1 of 75%

ACL 2012
Jeju, South Korea

Future work

Verb supersenses
Supersense annotation in other languages
Building an Arabic supersense tagger

www.ark.cs.cmu.edu/ArabicSST